

WHAT IS CLAIMED IS:

1. A speech recognition refinement system comprising:
 - a problematic word identifier that divides initial vocabulary words from
5 an initial speech recognition dictionary into problematic words
and non-problematic words according to pre-defined
identification criteria;
 - a candidate generator that analyzes said problematic words to produce
one or more pronunciation candidates for each of said
10 problematic words;
 - an optimization module that performs an optimization process for
refining said one or more pronunciation candidates according to
one or more optimization criteria, said optimization process
generating optimized problematic pronunciations; and
 - 15 a dictionary refinement manager that combines said optimized
problematic pronunciations with non-problematic pronunciations
of said non-problematic words to produce a refined speech
recognition dictionary for use by a speech recognition system.
- 20 2. The system of claim 1 wherein said pre-defined identification criteria
require that said problematic words each have a short-duration
characteristic, a common-use characteristic, and a high recognition-error
characteristic.
- 25 3. The system of claim 1 wherein said pre-defined identification criteria
include a short-duration characteristic which requires that said problematic
words each be spelled with fewer than five letters.

30

4. The system of claim 1 wherein said pre-defined identification criteria include a common-use characteristic that is quantified by analyzing an extensive training database of speech samples to determine which of said initial vocabulary words are frequently represented in said extensive training
5 database.

5. The system of claim 1 wherein said pre-defined identification criteria include a high recognition-error characteristic that is quantified by referring to a confusion matrix that includes separate recognition error rates for each
10 of said initial vocabulary words in said initial speech recognition dictionary when compared with all other of said initial vocabulary words in said initial speech recognition dictionary.

6. The system of claim 1 wherein said candidate generator includes a
15 phonetic recognizer that generates phone strings corresponding to said problematic words, said candidate generator also including a sequence analyzer that performs a multiple sequence analysis procedure upon said phone strings to produce said pronunciation candidates.

20 7. The system of claim 1 wherein said candidate generator includes a phonetic recognizer that sequentially processes speech data from multiple different utterances for each of said problematic words to produce individual phone strings that each represent corresponding intermediate pronunciations of respective ones of said problematic words.

25

30

8. The system of claim 1 wherein said candidate generator includes a sequence analyzer that perform one or more multiple sequence alignment procedures upon phone strings derived for each of said problematic words, said sequence analyzer aligning said phone strings, and then comparing
5 corresponding phones in each phone position of said phone strings to determine whether said corresponding phones indicate a consensus for identifying said pronunciation candidates.

9. The system of claim 1 wherein said optimization process supports
10 optimization functions that include selecting said pronunciation candidates, refining said pronunciation candidates, and deleting said pronunciation candidates.

10. The system of claim 1 wherein said optimization process supports
15 optimization functions that include selecting said pronunciation candidates for said refined speech recognition dictionary by choosing from consensus pronunciation candidates, majority pronunciation candidates, and plurality pronunciation candidates.

20 11. The system of claim 1 wherein said optimization process supports optimization functions that include refining said pronunciation candidates by adding or removing phones from said pronunciation candidates according to refinement criteria that include phonological rules of assimilation and coarticulation, physical limitations of human vocal tracks with regard to
25 producing certain phone sequences, contextual conditions such as inappropriate sequences of words and phones, and characteristics of dialectal and accent variations.

12. The system of claim 1 wherein said optimization process supports
30 optimization functions that include deleting said pronunciation candidates when recognition-error rates exceed a certain pre-defined threshold error level, said recognition-error rates being accessed from a confusion matrix.

13. The system of claim 1 wherein said optimization module automatically performs said optimization process according to said pre-defined optimization criteria by utilizing an expert system.

5

14. The system of claim 1 wherein said optimization module generates an optimization graphical user interface for a system user to interactively participate in said optimization process for converting said pronunciation candidates into said optimized problematic pronunciations.

10

15. The system of claim 14 wherein said optimization graphical user interface includes a word pane that displays selected entries from said initial speech recognition dictionary that match search criteria supplied by a system user, said word pane also displaying word lengths for said selected entries, pronunciations for said selected entries, total numbers of said initial vocabulary words from said speech recognition dictionary that share common pronunciations with said selected entries, and recognition accuracy rates for said selected entries provided from a confusion matrix.

20

16. The system of claim 14 wherein said optimization graphical user interface includes a candidate pane that displays a consensus pronunciation candidate, a majority pronunciation candidate, and a plurality pronunciation candidate for one of said initial vocabulary words selected from a word pane.

25

17. The system of claim 14 wherein said optimization graphical user interface includes a pronunciation pane that shows all initial pronunciations in said speech recognition dictionary for one of said initial vocabulary words selected from a word pane.

30

18. The system of claim 14 wherein said optimization graphical user interface includes a confusion pane that displays selected ones of said initial vocabulary words from said initial speech recognition dictionary that have an identical pronunciation that conflicts with a selected pronunciation from a pronunciation pane, said confusion pane also displaying corresponding respective recognition error rates.

19. The system of claim 1 wherein said problematic word identifier, said candidate generator, and said optimization module are implemented as part of said dictionary refinement manager, said dictionary refinement manager producing said refined speech recognition dictionary to improve speech recognition accuracy for spontaneous speech that includes certain spoken informalities which are incorporated into said optimized problematic pronunciations.

20. The system of claim 1 wherein said refined speech recognition dictionary is utilized by said speech recognition system during speech recognition procedures instead of using said initial speech recognition dictionary, said problematic word identifier, said candidate generator, said optimization module, and said dictionary refinement manager iteratively regenerating subsequent refined speech recognition dictionaries to further improve recognition accuracy rates of said speech recognition system for spontaneous speech.

21. A speech recognition refinement method comprising:
dividing initial vocabulary words from an initial speech recognition
dictionary into problematic words and non-problematic words
according to pre-defined identification criteria by utilizing a
5 problematic word identifier;
analyzing said problematic words with a candidate generator to
produce one or more pronunciation candidates for each of said
problematic words;
performing an optimization process with an optimization module to
10 refine said one or more pronunciation candidates according to
one or more optimization criteria, said optimization process
generating optimized problematic pronunciations; and
utilizing a dictionary refinement manager to combine said optimized
15 problematic pronunciations with non-problematic pronunciations
of said non-problematic words to produce a refined speech
recognition dictionary for use by a speech recognition system.
22. The method of claim 21 wherein said pre-defined identification criteria
require that said problematic words each have a short-duration
20 characteristic, a common-use characteristic, and a high recognition-error
characteristic.
23. The method of claim 21 wherein said pre-defined identification criteria
include a short-duration characteristic which requires that said problematic
25 words each be spelled with fewer than five letters.
24. The method of claim 21 wherein said pre-defined identification criteria
include a common-use characteristic that is quantified by analyzing an
extensive training database of speech samples to determine which of said
30 initial vocabulary words are frequently represented in said extensive training
database.

25. The method of claim 21 wherein said pre-defined identification criteria include a high recognition-error characteristic that is quantified by referring to a confusion matrix that includes separate recognition error rates for each of said initial vocabulary words in said initial speech recognition dictionary
5 when compared with all other of said initial vocabulary words in said initial speech recognition dictionary.

26. The method of claim 21 wherein said candidate generator includes a phonetic recognizer that generates phone strings corresponding to said
10 problematic words, said candidate generator also including a sequence analyzer that performs a multiple sequence analysis procedure upon said phone strings to produce said pronunciation candidates.

27. The method of claim 21 wherein said candidate generator includes a
15 phonetic recognizer that sequentially processes speech data from multiple different utterances for each of said problematic words to produce individual phone strings that each represent corresponding intermediate pronunciations of respective ones of said problematic words.

20 28. The method of claim 21 wherein said candidate generator includes a sequence analyzer that perform one or more multiple sequence alignment procedures upon phone strings derived for each of said problematic words, said sequence analyzer aligning said phone strings, and then comparing corresponding phones in each phone position of said phone strings to
25 determine whether said corresponding phones indicate a consensus for identifying said pronunciation candidates.

29. The method of claim 21 wherein said optimization process supports optimization functions that include selecting said pronunciation candidates,
30 refining said pronunciation candidates, and deleting said pronunciation candidates.

30. The method of claim 21 wherein said optimization process supports optimization functions that include selecting said pronunciation candidates for said refined speech recognition dictionary by choosing from consensus pronunciation candidates, majority pronunciation candidates, and plurality pronunciation candidates.

31. The method of claim 21 wherein said optimization process supports optimization functions that include refining said pronunciation candidates by adding or removing phones from said pronunciation candidates according to refinement criteria that include phonological rules of assimilation and coarticulation, physical limitations of human vocal tracks with regard to producing certain phone sequences, contextual conditions such as inappropriate sequences of words and phones, and characteristics of dialectal and accent variations.

32. The method of claim 21 wherein said optimization process supports optimization functions that include deleting said pronunciation candidates when recognition-error rates exceed a certain pre-defined threshold error level, said recognition-error rates being accessed from a confusion matrix.

33. The method of claim 21 wherein said optimization module automatically performs said optimization process according to said pre-defined optimization criteria by utilizing an expert system.

34. The method of claim 21 wherein said optimization module generates an optimization graphical user interface for a system user to interactively participate in said optimization process for converting said pronunciation candidates into said optimized problematic pronunciations.

35. The method of claim 34 wherein said optimization graphical user interface includes a word pane that displays selected entries from said initial speech recognition dictionary that match search criteria supplied by a system user, said word pane also displaying word lengths for said selected entries,
5 pronunciations for said selected entries, total numbers of said initial vocabulary words from said speech recognition dictionary that share common pronunciations with said selected entries, and recognition accuracy rates for said selected entries provided from a confusion matrix.

10 36. The method of claim 34 wherein said optimization graphical user interface includes a candidate pane that displays a consensus pronunciation candidate, a majority pronunciation candidate, and a plurality pronunciation candidate for one of said initial vocabulary words selected from a word pane.

15 37. The method of claim 34 wherein said optimization graphical user interface includes a pronunciation pane that shows all initial pronunciations in said speech recognition dictionary for one of said initial vocabulary words selected from a word pane.

20 38. The method of claim 34 wherein said optimization graphical user interface includes a confusion pane that displays selected ones of said initial vocabulary words from said initial speech recognition dictionary that have an identical pronunciation that conflicts with a selected pronunciation from a pronunciation pane, said confusion pane also displaying corresponding
25 respective recognition error rates.

39. The method of claim 21 wherein said problematic word identifier, said candidate generator, and said optimization module are implemented as part of said dictionary refinement manager, said dictionary refinement manager producing said refined speech recognition dictionary to improve speech
5 recognition accuracy for spontaneous speech that includes certain spoken informalities which are incorporated into said optimized problematic pronunciations.

40. The method of claim 21 wherein said refined speech recognition
10 dictionary is utilized by said speech recognition system during speech recognition procedures instead of using said initial speech recognition dictionary, said problematic word identifier, said candidate generator, said optimization module, and said dictionary refinement manager iteratively regenerating subsequent refined speech recognition dictionaries to further
15 improve recognition accuracy rates of said speech recognition system for spontaneous speech.

20

25

30

41. A computer-readable medium comprising program instructions for refining a speech recognition system by:

dividing initial vocabulary words from an initial speech recognition dictionary into problematic words and non-problematic words according to pre-defined identification criteria by utilizing a problematic word identifier;

analyzing said problematic words with a candidate generator to produce one or more pronunciation candidates for each of said problematic words;

performing an optimization process with an optimization module to refine said one or more pronunciation candidates according to one or more optimization criteria, said optimization process generating optimized problematic pronunciations; and

utilizing a dictionary refinement manager to combine said optimized problematic pronunciations with non-problematic pronunciations of said non-problematic words to produce a refined speech recognition dictionary for use by said speech recognition system.

42. A speech recognition refinement system comprising:
means for dividing initial vocabulary words from an initial speech
recognition dictionary into problematic words and non-
problematic words according to pre-defined identification criteria;
5 means for analyzing said problematic words to produce one or more
pronunciation candidates for each of said problematic words;
means for performing an optimization process to refine said one or
more pronunciation candidates according to optimization criteria,
said optimization process generating optimized problematic
10 pronunciations; and
means for combining said optimized problematic pronunciations with
non-problematic pronunciations of said non-problematic words
to produce a refined speech recognition dictionary for use by a
speech recognition system.

15 43. A speech recognition refinement system comprising:
a word identifier configured to identify problematic words and non-
problematic words according to identification criteria;
an optimization module configured to refine one or more pronunciation
20 candidates for each of said problematic words to produce
optimized problematic pronunciations; and
a refined speech recognition dictionary configured to include said
optimized problematic pronunciations and non-problematic
pronunciations of said non-problematic words.

44. A speech recognition refinement system comprising:
- a problematic word identifier that divides initial vocabulary words from an initial speech recognition dictionary into problematic words and non-problematic words according to pre-defined identification criteria which require that said problematic words each have a short duration, be commonly used, and exhibit a high likelihood of recognition error;
 - a candidate generator that analyzes said problematic words to produce one or more pronunciation candidates for each of said problematic words, said candidate generator including a phonetic recognizer that generates phone strings corresponding to said problematic words, said candidate generator also including a sequence analyzer that performs a multiple sequence analysis procedure upon said phone strings to identify said pronunciation candidates;
 - an optimization module that performs an optimization process for refining said one or more pronunciation candidates according to one or more optimization criteria, said optimization process generating optimized problematic pronunciations, said optimization process supporting optimization functions that include selecting said pronunciation candidates in an unchanged form, refining said pronunciation candidates, and deleting said pronunciation candidates; and
 - a dictionary refinement manager that combines said optimized problematic pronunciations with non-problematic pronunciations of said non-problematic words to produce a refined speech recognition dictionary for use by a speech recognition system.

45. A refined speech recognition dictionary implemented by:
dividing initial vocabulary words from an initial speech recognition
dictionary into problematic words and non-problematic words
according to pre-defined identification criteria;
5 analyzing said problematic words to produce one or more
pronunciation candidates for each of said problematic words;
performing an optimization process to refine said one or more
pronunciation candidates according to one or more optimization
criteria, said optimization process generating optimized
10 problematic pronunciations; and
combining said optimized problematic pronunciations with non-
problematic pronunciations of said non-problematic words to
produce said refined speech recognition dictionary for use by a
speech recognition system.

15